Investigating the Relationship between Topology and Evolution in a Dynamic Nematode Odor Genetic Network

David A. Fitzpatrick and Damien M. O’Halloran

1 Genome Evolution Laboratory, Department of Biology, National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland
2 Department of Biological Sciences, The George Washington University, 333 Lisner Hall, 2023 G Street NW, Washington, DC 20052, USA
3 Institute for Neuroscience, The George Washington University, 636 Ross Hall, 2300 I Street NW, Washington, DC 20037, USA

Correspondence should be addressed to Damien M. O’Halloran, damienoh@gwu.edu

Received 17 May 2012; Revised 6 August 2012; Accepted 29 August 2012

Academic Editor: Amitabh Joshi

Copyright © 2012 D. A. Fitzpatrick and D. M. O’Halloran. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The relationship between biological network architectures and evolution is unclear. Within the phylum nematoda olfaction represents a critical survival tool. For nematodes, olfaction contributes to multiple processes including the finding of food, hosts, and reproductive partners, in the making of developmental decisions, and to the evasion from predators. Studies into the olfactory system of the model nematode Caenorhabditis elegans have yielded detailed descriptions of the molecular and cellular pathways that subserve the olfactory system [1–4]. These signaling pathways appear highly conserved across very divergent nematode species, and the sensory neurons have clear anatomical orthologs in distantly related nematodes [5]. Within the olfactory system of C. elegans, the ability to detect dilute volatile odors is mostly conferred by three pairs of neurons termed the Amphid Wing cells type A (AWA), Amphid Wing cells type B (AWB), and the Amphid Wing cells type C (AWC) [3, 4]. These cells are primary sensory neurons located within the sensory amphid organ of the head that forms part of an anatomically distinct subclass of amphid neurons in that they do not extend processes through the amphid opening, but instead their distal ciliated endings terminate inside a glial sheath cell [6]. Here we describe a composite odor genetic network in C. elegans that encompasses all three pairs of volatile odor-detecting neurons. We used this network to identify orthologous genes of the odor network in the nematode Pristionchus pacificus.

We chose the nematode P. pacificus based upon three criteria: (1) it exhibits a similar lifestyle to that of C. elegans, in that they are both self-fertilizing hermaphrodites that will feed on bacteria, (2) conservation of the olfactory signaling pathway of P. pacificus to C. elegans has been validated experimentally [7, 8], and (3) because it is sufficiently divergent to generate meaningful divergence data (estimated divergence between Caenorhabditis and Pristionchus is between 280 and 430 million years ago [9, 10]). One of our goals in this study was to compare selective pressures across the nodes of the odor network by comparing substitutions at silent sites to that of nonsilent sites; this analysis is best done using more divergent sequences [11]. Using this nematode odor network we searched for relationships between pathway position with divergence, diversity, and contribution across the network. Then, we designed an interaction map of the network.

1. Introduction

For nematodes, olfaction is a central mode of survival. Olfaction contributes to the finding of food, hosts, reproductive partners, in the making of developmental decisions, and to the evasion from predators. Studies into the olfactory system of the model nematode Caenorhabditis elegans have yielded detailed descriptions of the molecular and cellular pathways that subserve the olfactory system [1–4]. These signaling pathways appear highly conserved across very divergent nematode species, and the sensory neurons have clear anatomical orthologs in distantly related nematodes [5]. Within the olfactory system of C. elegans, the ability to detect dilute volatile odors is mostly conferred by three pairs of neurons termed the Amphid Wing cells type A (AWA), Amphid Wing cells type B (AWB), and the Amphid Wing cells type C (AWC) [3, 4]. These cells are primary sensory neurons located within the sensory amphid organ of the head that forms part of an anatomically distinct subclass of amphid neurons in that they do not extend processes through the amphid opening, but instead their distal ciliated endings terminate inside a glial sheath cell [6]. Here we describe a composite odor genetic network in C. elegans that encompasses all three pairs of volatile odor-detecting neurons. We used this network to identify orthologous genes of the odor network in the nematode Pristionchus pacificus.

We chose the nematode P. pacificus based upon three criteria: (1) it exhibits a similar lifestyle to that of C. elegans, in that they are both self-fertilizing hermaphrodites that will feed on bacteria, (2) conservation of the olfactory signaling pathway of P. pacificus to C. elegans has been validated experimentally [7, 8], and (3) because it is sufficiently divergent to generate meaningful divergence data (estimated divergence between Caenorhabditis and Pristionchus is between 280 and 430 million years ago [9, 10]). One of our goals in this study was to compare selective pressures across the nodes of the odor network by comparing substitutions at silent sites to that of nonsilent sites; this analysis is best done using more divergent sequences [11]. Using this nematode odor network we searched for relationships between pathway position with divergence, diversity, and contribution across the network. Then, we designed an interaction map of the network.
We only included orthologs that are therefore not possible to reconstruct phylogeny (see Figures S1–S7).

Over novel nodes within the network.

We design a framework that facilitates a level of predictive power over novel nodes within the network.

2. Materials and Methods

2.1. Sequences. Pristionchus pacificus orthologs were located by cross-referencing matches using the orthology databases: InParanoid_7 [12], OrthoMCL database [13], and the OMA orthology matrix Browser [14]. For each node in our network WormBase (version WS231) has defined a curated ortholog using WormBase-Compara [15]; however, we corroborated these predictions using reciprocal Blast searches, and, by inferring relationships by reconstructing phylogenetic trees (see Figures S1–S7).

2.2. Analysis of Molecular Data. Synonymous (dS) and nonsynonymous (dN) substitution rates for orthologs were estimated using the methods of Yang and Nielsen [19] as implemented in yn00 in the PAML suite [20]. To test the null hypothesis that there is no above average selective pressure on these genes, we performed a randomization test where we determined the average dN/dS value for 50,000 randomly assembled networks and compared with the average dN/dS value for our odor network. Random networks were 16 genes in size and sampling permitted replacement. Measures of nucleotide diversity (π) were performed using DnaSP version 5 [21].

2.3. Network Analysis. Selection of vertices for the odor network was determined by mining literature databases. Network analysis was calculated using Cytoscape version 2.8.2

<table>
<thead>
<tr>
<th>C. elegans gene</th>
<th>P. pacificus ortholog</th>
<th>Evidence of orthology</th>
</tr>
</thead>
<tbody>
<tr>
<td>goa-1</td>
<td>PPA23648/Ppa-goa-1</td>
<td>WormBase-Compara</td>
</tr>
<tr>
<td>odr-3</td>
<td>PPA14189/Ppa-odr-3</td>
<td>InParanoid; OMA; WormBase-Compara</td>
</tr>
<tr>
<td>egl-30</td>
<td>PPA31690/Ppa-egl-30</td>
<td>WormBase-Compara</td>
</tr>
<tr>
<td>gpa-3</td>
<td>PPA09867/Ppa-gpa-3</td>
<td>InParanoid; OMA; WormBase-Compara</td>
</tr>
<tr>
<td>gpa-5</td>
<td>PPA10789</td>
<td>OMA; WormBase-Compara</td>
</tr>
<tr>
<td>gpa-13</td>
<td>PPA31402/Ppa-gpa-13</td>
<td>WormBase-Compara</td>
</tr>
<tr>
<td>eat-4</td>
<td>PPA15025/Ppa-eat-4</td>
<td>InParanoid; WormBase-Compara</td>
</tr>
<tr>
<td>dkg-1</td>
<td>PPA00312/Ppa-dkg-1</td>
<td>InParanoid; OMA; WormBase-Compara</td>
</tr>
<tr>
<td>odr-1</td>
<td>PPA17708</td>
<td>InParanoid; OMA</td>
</tr>
<tr>
<td>daf-11</td>
<td>PPA14907/Ppa-daf-11</td>
<td>WormBase-Compara</td>
</tr>
<tr>
<td>tax-6</td>
<td>PPA09320/Ppa-tax-6</td>
<td>WormBase-Compara</td>
</tr>
<tr>
<td>pkg-1</td>
<td>PPA27475/Ppa-pkg-1</td>
<td>InParanoid; OMA; WormBase-Compara</td>
</tr>
<tr>
<td>tax-2</td>
<td>PPA07436/Ppa-tax-2</td>
<td>InParanoid; OMA; WormBase-Compara</td>
</tr>
<tr>
<td>tax-4</td>
<td>PPA02388/Ppa-tax-4</td>
<td>InParanoid; OMA; WormBase-Compara</td>
</tr>
<tr>
<td>rgs-3</td>
<td>PPA01970/Ppa-rgs-3</td>
<td>InParanoid; OMA; WormBase-Compara</td>
</tr>
<tr>
<td>arr-1</td>
<td>PPA21763/Ppa-arr-1</td>
<td>WormBase-Compara</td>
</tr>
</tbody>
</table>

1Relationship support was confirmed by reconstructing phylogenetic trees for each gene except rgs-3 and arr-1 as they represent single-gene families, and therefore not possible to reconstruct phylogeny (see Figures S1–S7).

and investigated relationships between various metrics of interaction with molecular evolution, and contribution across the network. Our data is not a definitive map of odor signaling in C. elegans but represents a snap shot of current data, and by uncovering robust associations between network topology, evolution, and function we may ultimately design a framework that facilitates a level of predictive power over novel nodes within the network.

Table 1: Orthology support for each node within the odor-signaling network.
The network was treated as undirected and all network analyses available through Cytoscape version 2.8.2 were examined; these are average shortest path length, betweenness centrality, closeness centrality, clustering coefficient, degree, eccentricity, neighborhood connectivity, radiality, stress, and topological coefficient. The degree \( k_n \) of a node \( n \) is defined as the number of edges linked to \( n \). The clustering coefficient \( (C_n) \) reveals how connected the neighborhood of a node is by calculating the fraction of neighboring pairs, and for a node \( n \) it is defined as

\[
C_n = \frac{2e_n}{k_n(k_n-1)},
\]

(1)

where \( k_n \) is the degree of \( n \) and \( e_n \) is the number of connected pairs between all neighbors of \( n \). Betweenness centrality \( (C_b) \) is a measure of the fraction of shortest paths between node pairs \( (s, t) \) that pass through the node of interest, and for the node \( n \) it is calculated using the following formula:

\[
C_b(n) = \sum_{s \neq n \neq t} \frac{\sigma_{st}(n)}{\sigma_{st}},
\]

(2)

where \( s \) and \( t \) are nodes different from \( n \), \( \sigma_{st}(n) \) denotes the number of shortest paths from \( s \) to \( t \), and \( \sigma_{st}(n) \) is the number of shortest paths from \( s \) to \( t \) on which \( n \) lies.

All chemotaxis indices (for both wildtype and mutant animals) were mined from previous publications (see references below). The chemotaxis index difference \( (C.I._{diff}) \) for each mutant within each neuron (AWA, AWB, and AWC) was determined by calculating the difference between the wildtype chemotaxis index \( (C.I._{wt}) \) and mutant chemotaxis index \( (C.I._{mut}) \). For example, a mutant that presents an average AWA chemotaxis defect \( \text{C.I.} = 0.5 \), compared to the wildtype C.I. \( (C.I._{wt-AWA}) \) of 0.9, would have a \( C.I._{diff-AWA} = 0.4 \) (i.e., subtract 0.5 from 0.9). In the case of long-term adaptation (LTA) mutants the \( C.I._{diff} \) was calculated by subtracting the C.I. value for LTA from \( C.I._{wt} \) of unadapted animals. Wildtype odor: 0.9 (\( C.I._{wt-AWA} \)); 0.85 (\( C.I._{wt-AWC} \))[4, 23]. Wildtype long-term adaptation: 0.5 (\( C.I._{diff-AWA} \)); 0.65 (\( C.I._{diff-AWC} \))[24, 25].

\[\text{odr-3}: \ 0.6 \ (C.I._{diff-AWA}); \ -0.7 \ (C.I._{diff-AWB}); \ 0.5 \ (C.I._{diff-AWC})[23, 26]. \]

\[\text{r5s-3}: \ 0.85 \ (C.I._{diff-AWC})[27]. \text{daf-11}: \ 0.6 \ (C.I._{diff-AWB}); \ 0.5 \ (C.I._{diff-AWC})[28]. \text{dkg-1}: \ 0 \ (C.I._{diff-AWC})[29]. \text{eat-4}: \ 0.55 \ (C.I._{diff-AWC})[30]. \text{egl-30}: \ 0.65 \ (C.I._{diff-AWC})[29]. \]

\[\text{goa-1}: \ 0.6 \ (C.I._{diff-AWA}); \ 0.6 \ (C.I._{diff-AWC})[29]. \text{odr-1}: \ 0 \ (C.I._{diff-AWA}); \ 0.85 \ (C.I._{diff-AWC})[31, 32]. \text{tax-2}: \ 0.2 \ (C.I._{diff-AWA}); \ -0.8 \ (C.I._{diff-AWB}); \ 0.75 \ (C.I._{diff-AWC})[23, 33], \text{tax-4}: \ 0.5 \ (C.I._{diff-AWA}); \ 0.75 \ (C.I._{diff-AWC})[23, 34]. \text{egl-4}: \ 0.8 \ (C.I._{diff-AWA}); \ 0.65 \ (C.I._{diff-AWC})[25, 34]. \]

The phenotype index (PI) was then calculated as follows:

\[
P.I. = \sqrt{\left(\frac{C.I._{diff-mut-AWA}}{C.I._{wt-AWA}}\right)^2 + \left(\frac{C.I._{diff-mut-AWB}}{C.I._{wt-AWB}}\right)^2 + \left(\frac{C.I._{diff-mut-AWC}}{C.I._{wt-AWC}}\right)^2}. \]

(3)

The PI for LTA mutants was measured by calculating the relative contribution of the LTA mutant to the adaptation response. For example, if the LTA response of wildtype animals normally decreases the C.I. value to 0.3 from 0.9, then the \( C.I._{diff} = 0.6 \); if the mutant is fully required (i.e., there is no LTA response at all in the mutant) for the LTA response, then the PI is calculated as follows: \( 0.6/0.6 = 1 \); which reflects the total contribution of the gene product to LTA in wildtype animals but that which is defective in the LTA mutant. The weighted phenotype index (WPI) was calculated for each group of mutants as follows:

\[
\sum_{i=1}^{n} \left(\frac{P.I.}{mut^x}(\text{mut}^x)\right) \frac{w_{mut^x}}{(\text{mut}^x)} \]

(4)

where \( P.I. \) is the phenotype index (P.I.) for the mutant \( x \) (\( \text{mut}^x \)), and \( w_{mut^x} \) represents the number of odors for which mutant \( x \) (\( \text{mut}^x \)) exhibits a mutant phenotype.

3. Results

To compare divergence rate with pathway position in our network of dynamic odor transduction, we identified orthologous genes of the C. elegans odor network in P. pacificus, and divided the odor-signaling pathway members (Figure 1(a)) within this network into 3 categories: (I) regulators Class 1 (GPA-3 [35, 36], ODR-3 [26], GPA-5 [35, 36], GPA-13 [35, 36], ARR-1 [37] and RGS-3 [27]); (II) regulators Class 2 (EGL-4 [25], TAX-6 [38], ODR-1 [31], DAF-11 [28], TAX-2 [33], and TAX-4 [39]); (III) actuators (GOA-1 [40], EGL-30 [40], DGK-1 [40], and EAT-4 [41]). After binning pathway members into each category, we then calculated divergence within each group by comparing substitutions at silent sites to that of nonsilent sites and looked for patterns of correlation between topology and evolutionary rate. We did not observe a significant association between pathway position and divergence (pathway position: Spearman rank-order correlation coefficient \( r_s = 0.37, P = 0.15 \), Kendall’s \( \tau = 0.31, P = 0.13 \)). We also designed an orthologous intragenus network using Caenorhabditis japonica and compared substitutions at silent sites to that of non-silent sites and looked for a pattern of correlation with pathway position. Using this intragenus network we also did not observe significant association between pathway position and divergence (pathway position: \( r_s = 0.28, P = 0.3 \), Kendall’s \( \tau = 0.17, P = 0.4 \)), suggesting that the pathway position alone may not shape constraint in our odor network.

Next, we designed an interaction map of the network and tested for correlations between divergence and various topological metrics (Figure 1(b)). The interaction-based odor network is a composite map of the pathways within the three pairs of neurons that subserves volatile odor recognition.
Figure 1: Odor-signaling pathway in *C. elegans* illustrated by pathway position and based on genetic interactions. (a) Summary of volatile odor signaling in *C. elegans*. (b) Schematic of an interaction-based odor network from *C. elegans*.

Figure 2: Plotting the relationship between divergence with measures of betweenness centrality or degree, and the relationship between contribution with measures of betweenness centrality or degree in a dynamic nematode odor network. (a) Divergence (left y-axis in blue) plotted against increasing measures of betweenness centrality (x-axis), and contribution (right y-axis in red) plotted against increasing measures of betweenness centrality. (b) Divergence (left y-axis in blue) plotted against increasing measures of degree (x-axis), and contribution (right y-axis in red) plotted against increasing measures of degree.
in *C. elegans* (these cells are the AWA, AWB, and AWC). The interaction network was designed based upon genetic interactions and localization studies. Edges may represent interactions from only one of the three odor sensing pairs of neurons, but placed in a composite framework of all three pairs of odor-sensing neurons to obtain an informative number of vertices. The network was treated as undirected. The network comprised 16 nodes with a network diameter of 5, a characteristic path length of 2.825, and an average number of neighbors equal to 2. For the interaction network we observed that a strong correlation between number of neighbors and average neighborhood connectivity (correlation = 0.857; coefficient of determination $R^2 = 0.648$). Plotting the number of neighbors against average clustering coefficient yields a strong pattern of correlation using a linear function (correlation = 0.868; coefficient of determination $R^2 = 0.753$); we observed that a decrease in clustering coefficient with an increase in number of neighbors as is typical of many networks. Network analysis was performed and the output of the analyses was tested for correlations with divergence by calculating the evolutionary rate for each node by comparing substitutions at silent sites to that of non-silent divergence when organized by pathway position, betweenness centrality, and degree. In each case we did not observe a significant pattern of correlation (betweenness: $r_s = -0.35, P = 0.06$, Kendall’s $\tau = -0.3, P = 0.13$; Pathway position: $r_s = 0.4, P = 0.08$, Kendall’s $\tau = 0.37, P = 0.06$; degree: $r_s = -0.47, P = 0.06$, Kendall’s $\tau = -0.38, P = 0.06$).

Next, we developed a metric of contribution within the network we termed the weighted phenotype index (WPI), and compared the total contribution of each group with divergence when organized by pathway position, betweenness, or degree. In the case of pathway position we observed a weak correlation between divergence and contribution ($r = 0.44$), and in the case of betweenness or degree we observed strong negative correlations with contribution and divergence (Figure 2(a) betweenness: $r = -0.92$; Figure 2(b) degree: $r = -0.76$). We also examined the correlation between quantities of betweenness or degree for each node in our network with contribution, and again we observed robust correlations (Figure 2(a) betweenness: $r = 0.9$; Figure 2(b) degree: $r = 0.79$). From this analysis we uncovered patterns of association between measures of betweenness, and degree with evolutionary rate and contribution.

The most highly connected nodes within our network are that of *odr-3* and *egl-4* (also called *pkg-1*). *ODR-3* is a Gα subunit protein that transduces multiple stimuli within the network [26]. *EGL-4* is a protein kinase G that facilitates both primary signal transduction as well as desensitization responses within the network [25, 34]. To
examine how molecular diversity varies across these hub vertices, we performed a sliding window examination of the coding region of each gene using 100-base-pair windows in increments of 20 base pairs. In the case of egl-4, we observed only a few polymorphic peaks, with each peak associating with areas between domains of functional importance (Figure 3(a)); these are the N-terminal low-complexity domain (LC), the coiled-coil domain (CC), two cGMP-binding domains (cNMP), and the serine/threonine kinase domain (S/T K). In the case of odr-3, we observed more variability overall compared with that of egl-4, but again with most variable peaks associating with areas between domains of functional significance (Figure 3(b)); in particular, the five alpha helices (G1–G5) that comprise the GTPase domain, the receptor interaction C-terminal (CT) as well as the fatty acid modification site (M) (Figure 3(b)). We observed similar scales for nucleotide diversity for both egl-4 (π = 0.27) and odr-3 (π = 0.25) with each gene exhibiting polymorphic peaks in areas between domains of functional importance. The similar pattern of diversity is in keeping with similar trends of divergence between egl-4 and odr-3: egl-4 dS/dN = 0.0041, dN = 0.1183, dS = 29.1777; odr-3 dS/dN = 0.0054, dN = 0.1132, and dS = 21.1204.

Overall, nodes within the odor network are undergoing purifying selection, with an average dS/dN value across the network of 0.012 for P. pacificus versus C. elegans. To place this value in a context of global divergence rates between C. elegans and P. pacificus we examined divergence across 5,666 pairs of 1 : 1 orthologs, then generated randomized data sets (50,000 in total) comprising an equal number of nodes as our odor network (i.e., 16), and examined the frequency of mean dS/dN values in each case (Figure 4, blue bars). From this analysis we found an average dS/dN value = 0.14 which is more than an order of magnitude higher (~11.6X) than what we observed for our dynamic odor network. To control for the number of Gα subunit proteins within our odor network we also conducted a randomization study whereby dS/dN averages from 50,000 random data sets of 16 nodes that contain 6 Gα subunit genes in each case were generated (Figure 4, red bars). In this control experiment we still observed a significantly higher average dS/dN value for the odor network (average = 0.106) than for our odor network, taken together, this suggests that the dynamic odor-signaling network is shaped by a more pronounced purifying selection than what guides global constraint across the genome.

4. Discussion

Positional rate variation (PRV) has been tested within numerous biosynthetic pathways where it has been demonstrated that upstream genes (perhaps on account of greater pleiotropy) undergo more selective constraint than downstream genes [42–44]. However, it is unlikely that this effect holds true for all signaling pathways as input and network architecture will vary widely. In the case of sensory signaling pathways that subserve an environment versus perception “arms race,” the idea of upstream genes exhibiting greater diversity may be more applicable. This trend was observed in a study of the Drosophila innate immune system where the authors demonstrated that recognition genes undergo higher levels of positive selection than immune effector genes [45]. In our network we did not observe a significant correlation between pathway position and divergence, which was in keeping with previous observations from data on intragenus networks of odor signaling in Caenorhabditis [46]. However, by comparing measures of degree or centrality with divergence, we did observe significant negative correlations. This may suggest that pathway position is not the only factor in shaping constraint or change in our network, but rather network characteristics such as connectivity may play major roles in guiding molecular evolution. Negative correlations between divergence with node relevance, and positive correlations between essentiality with node relevance have been reported previously and may represent a general trend of many biological networks [47, 48].

By comparing global levels of divergence between C. elegans and P. pacificus we found a pervasive theme of constraint across the genome; however, through our randomization study (Figure 4) we found that global levels of constraint are an order of magnitude lower than the level of constraint across the odor network. This suggests that the odor-signaling pathway represents a fixed circuit whose network properties are preserved by strong purifying selection. Low levels of divergence have been reported within Caenorhabditis for members of the olfactory pathway [46], and previously we have found that a large component in Regulators Class 1 and 2 has undergone extensive nematode-specific gene duplication events, namely, the Gα subunit proteins and guanylyl cyclase proteins [49, 50]. Many of

**Figure 4:** Frequency distribution of average dS/dN values from 50,000 randomization sets comprising 16 node groupings from 5,666 1 : 1 orthologs between C. elegans and P. pacificus (blue bars). Red bars indicate frequency distribution of average dS/dN values from 50,000 randomization sets of 16 node groupings that include 6 Gα subunit proteins in each; this control group was examined to investigate potential bias in our network by having 6 Gα subunit proteins within our odor genetic network. The x-axis represents binned divergence categories plotted against the frequency on the y-axis. The green arrow indicates the actual average dS/dN value of the nematode odor network.
these genes are either exclusively or highly expressed in primary sensory neurons [26, 28, 31, 35, 36, 51]. This genetic expansion facilitates multiple capacities that shape developmental and survival strategies through intercellular and intracellular processing of polymodal sensory input. Many of these duplicates are present in all four major clades of the phylum nematoda [49] suggesting that the odor-signaling network arose early in nematode evolution and has undergone neo-functionalization events that are preserved by strong functional constraint.

Acknowledgments

The authors are very grateful to Rolf Sommer and Christian Rodelsperger for providing them with Pristionchus pacificus coding sequence files and to Scott Hamilton for critical reading of this manuscript. They also wish to thank their two anonymous reviewers for their excellent suggestions.

References


Submit your manuscripts at http://www.hindawi.com